**HW 3**

**Part 2: Forecasting Unemployment using ARIMA model**

One of the most watched macroeconomic variables in the United States is the unemployment rate, the percentage of the workforce that is unemployed. A variety of federal, state, and local agencies track unemployment. An increase in unemployment anticipates greater demand for services from the government. Companies also watch this rate. Managers at a chain of fast-food restaurants, for instance, worry that rising levels of employment could produce pressure to raise wages, cutting into profits. **Your task is to build an ARIMA model to forecast unemployment rate.**

**Data source:**

**Series 1:** Annual unemployment Rate in California:

[**https://fred.stlouisfed.org/series/LAUST060000000000003A#0**](https://fred.stlouisfed.org/series/LAUST060000000000003A#0)

**Series 2:** Monthly unemployment Rate in the US:

[**https://fred.stlouisfed.org/series/UNRATENSA**](https://fred.stlouisfed.org/series/UNRATENSA)

Remarks:

* assume current time stamp = February of 2020. We will not use actual data after February 2020!
* Here is an applied article on unemployment forecasting using ARIMA models that can help review some concepts:<http://www.jaqm.ro/issues/volume-3,issue-2/pdfs/dobre_alexandru.pdf>

**Questions:**

1. Import both time series directly from FRED into python.
2. Visualize both series and interpret the patterns you see in data.
3. Create the acf plot for each series and describe the dependence structure in both of the original series.
4. Based on the visualizations you created in previous questions, do you think series are stationary or not? Why do we care whether data is stationary or not?
5. Carry out the stationarity tests. Do tests suggest the data is stationary?
6. What corrections would you recommend if the series is non-stationary?
7. Perform necessary preprocessing on both series (for example, differencing transformations if needed) and identify an appropriate ARIMA model(s) for each series.
8. What is the simplest models that provide an adequate description of the observed data (this is sometimes known as the principle of parsimony). Discuss with reference to residual plots, residual autocorrelations, and the Box test.
9. Select the best model using AIC and generate forecasts for the next (Future: March 2020,….) 12 months for monthly series (US series) and the next 3 years (Future: 2020,2021,2022) for annual series (CA series). Assume current time stamp = February of 2020
10. Plot each series (US and CA on separate graphs) and overlay fitted values and future forecasts.
11. Report MAPE (on historical period is enough) of the champion model for each series.
12. What is the best model using the auto arima function for each series?
13. Are the best auto models adequate? Discuss with reference to residual plots, residual autocorrelations, and the Box test.
14. Report MAPE (on historical period is enough) of the best auto arima model for each series.